STUDIA PHONOLOGICA XVII (1983)

Plosive Discrimination by Running Spectra in 40-ms Initial Segment

Shigeyoshi KITAZAWA and Shuji DOSHITA

Abstract

The present paper deals with invariant features for place of articulation. This paper describes results of consonant discrimination using 5 running spectra extracted during burst after 40-ms waveform of Japanese plosive-vowel syllables. The number of talkers analyzed was 48 and the phonetic environments were /p, t, k, b, d, g/ and five Japanese vowels. The features were extracted so as to be invariant with respect to both talker and vowel: First, a set of canonical variables was selected and transformed from spectrum variables in each frame, and then these canonical variables were integrated and used for final discrimination. Due to integration of consonantal features in these successive spectra, correct classification score was higher than that of a single spectrum experiment. We examined the feature distribution in the burst and the following transition spectrum, and the differences between voiced and voiceless plosives.

Introduction

The acoustic features for discrimination of place of articulation in stop consonants are not unique. The roles of the burst spectrum (Winitz et al., 1972) and the second- and third-formant loci (Delattre et al., 1955) as place cues were studied and understood separately.

Perception experiments with natural speech as well as synthetic speech stimuli have shown that listeners can identify place of articulation in stimuli containing only the release burst and the first few tens of milliseconds of glottal pulsing. Although performance of subjects improves as longer durations of the stimuli are presented (Blumstein and Stevens, 1980; Kewley-Port, 1983), performance is well above chance even with stimuli as short as 20- or so ms in duration.

It is widely said that there is acoustic invariance for place of articulation in initial stop consonants. Since the articulatory gesture for a given consonant is assumed to be relatively fixed, the corresponding acoustic cues for place should be invariant. Spectral analysis of the onset characteristics of natural speech utterances

Shigeyoshi KITAZAWA (北沢茂良) : Assistant, Department of Information Science, Kyoto University.
Shuji DOSHITA (堂下修司) : Professor, Department of Information Science, Kyoto University.

has indicated that distinguishing properties for place of articulation are found independently of the particular vowel context in which the utterances occur. These results were obtained using a relatively static representation based on the gross shape of the spectrum in a single time frame of analysis of 25.6-ms duration (Blumstein and Stevens, 1979), as well as by using a time-varying representation looking at the spectrum over longer time frames (Searle et al., 1979; Searle et al., 1980; Kewley-Port, 1983; Ide et al., 1983).

The work of Kewley-Port (1980) provided an alternative definition of the information-bearing characteristics of the waveform at consonant release. Specifically, she argued that the gross shape of the onset spectrum must be an insufficient cue for stop consonant place, because it does not incorporate the dynamic changes in spectral shape that occur during consonant release and subsequent articulatory movement toward the following vowel. She proposed three dynamic spectral features as invariant cues for stop consonant place. The tilt of spectrum at burst onset cues bilabial place if it is falling, and alveolar place if it is flat or rising. Late onset of low frequency energy functions as a cue for velar place, since velar stops are characterized by longer VOT that are bilabials or alveolars. Finally the presence of midfrequency peaks extending over time is a cue for velar place; these peaks reflect the resonant characteristics of the cavity anterior to the velar constriction. All of these features may be evaluated during the first 20-40-ms of the stop consonant-vowel waveform.

Regardless of the actual nature of the invariant cue for stop consonant place, the data of both Stevens and Blumstein (1978) and Kewley-Port (1980) indicate that apparently this cue resides in the first 20-40-ms of the syllable waveform.

The approach taken in this paper follows the concept that plosive consonants are characterized by several acoustic features: transitions, bursts, and timing, therefore feature integration improves recognition accuracy. If this is the case, then spectral analysis of first 20-40-ms of stop waveform provides effective features for place discrimination, possibly invariant for vowel context and speaker difference.

This paper extends the previous study in which 28 talkers' voiceless stops were investigated (Kitazawa, 1982). The number of talkers was increased and voiced stops were examined as well as voiceless ones. This paper describes further analysis exploiting time-varying spectral information as well feature distribution in the burst and following transition.

I. METHOD

A. Acoustic Segmentation of Stop-Consonant-Vowel Syllables

Acoustic analysis of plosive-vowel syllables reveals five quantitatively distinct segments before a stable vowel reached: (1) a period of occlusion (usually silent, though occasionally voiced); (2) a transient explosion (usually less than 20-ms) pro-

duced by shock excitation of the vocal tract upon release of occlusion; (3) a very brief (0–10–ms) period of frication, as articulators separate and air is blown through a narrow (though widening) constriction, as in the homorganic fricative; (4) a brief period (2–20–ms) of aspiration, within which may be detected noise–excited formant transitions, reflecting shifts in vocal–tract resonances as the main body of the tongue moves toward a position appropriate for the following vowel; (5) voiced formant transitions reflecting the final steps of tongue movement into the vowel position during the first few cycles of laryngeal vibration.

The present paper attempts to investigate acoustic features for place of articulation over these segments, mainly the burst portion comprising the explosion and aspiration. In this experiment, as in earlier studies of stop consonants, the burst frame was located visually in the waveform by the experimenter.

Voiced Japanese stops lack the aspiration phase. A short frication segment can be seen in /d/ and /g/. The duration of the /g/ and /k/ transients is longer than in any other stops. Uninterrupted voicing can be superimposed during all phases of /b/, /d/, and /g/.

The duration of the interval between the stop–release and the onset of voicing is refered to as Voice–Onset–Time (VOT) by Lisker and Abramson (1964). Since the VOT has important implication in the following discussion, Fig. 1 shows histograms of VOT for each consonant, which in the present experiment are visually measured using an interactive graphic terminal.

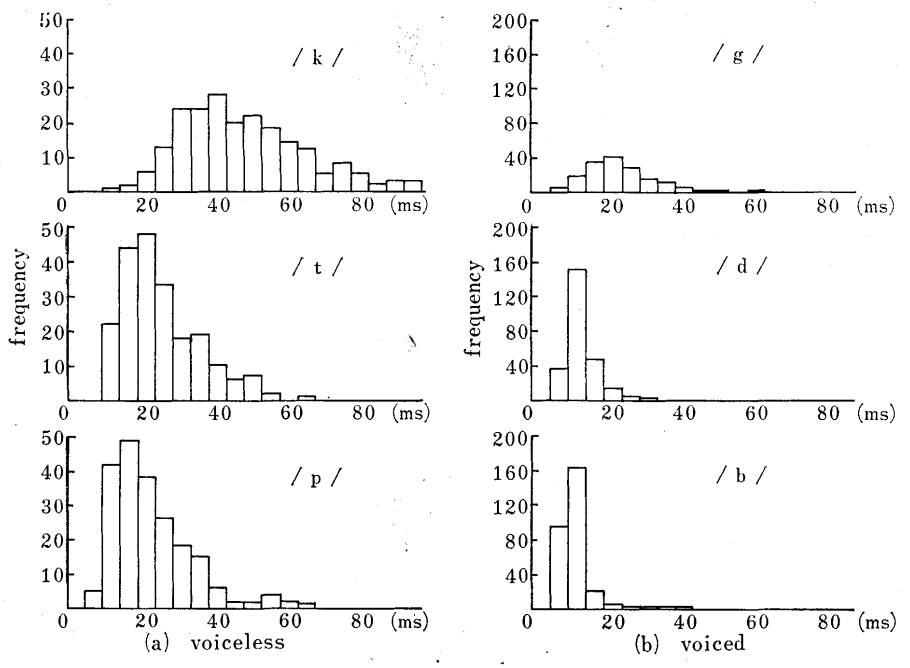

Fig. 1. Histograms of Voice–Onset–Time.

B. Subjects

Fourty-eight male speakers produced the set of consonant-vowel syllables that were analyzed in this study. Five of the speakers made three repetitions. For some speakers a few syllables were missing, however the number of repetions of a given syllable was less than 43. Only one of each syllable /p, t, k, b, d, g/ paired with/a, i, u, e, o/ was used in this study. Each syllable was read from ordered list. Further recording procedure is described in the previous paper (Kitazawa, 1982). Each syllable was digitized for analysis. Waveforms were first low-pass filtered at 8.9-kHz and then sampled at 18.5-kHz using a 12-bit analog to digital converter. The total number of syllables examined in this experiment was 869 from voiceless stops, and 837 from voiced stops.

C. Analysis of Running Spectrum

The waveforms were edited and differentiated once (pre-emphasized with coefficient 0.95). Linear prediction coefficients were calculated for each window using the autocorrelation method where a 25-ms window was used. Smoothed spectra were calculated by means of a discrete Fourier transform of the coefficients with added

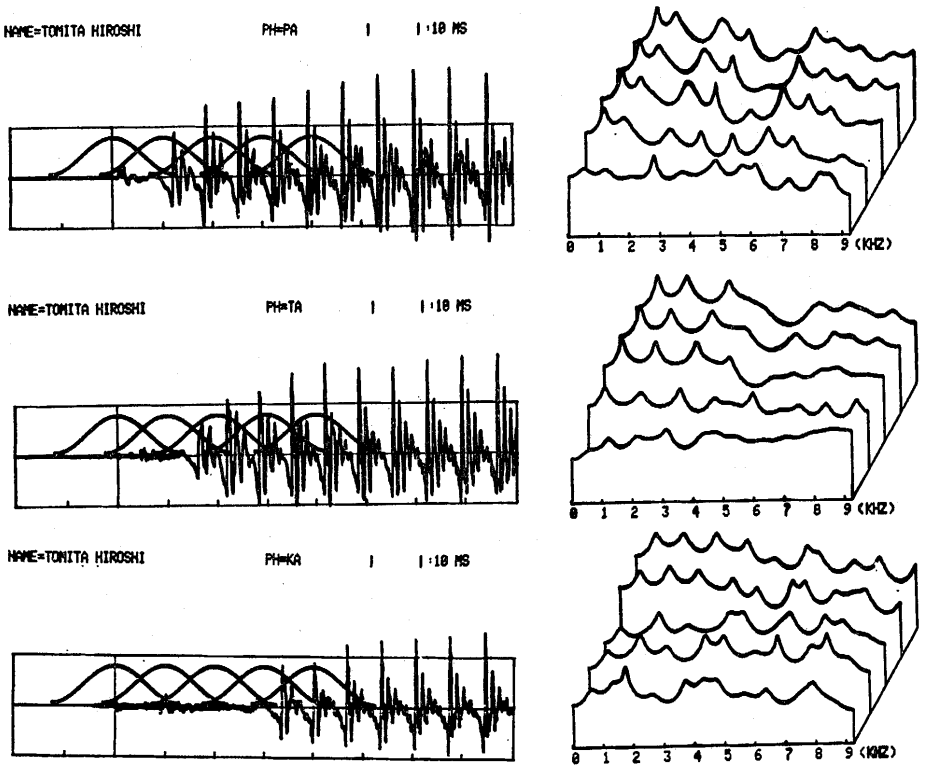

Fig. 2. Running spectra and analysis window. Window is 25-ms Hamming, and frame interval is 10-ms. The first window centered at the burst onset. The running spectra display transition from burst to target vowel /a/.

zeros. The resulting 256-point spectrum excluding the gain factor is a zero-mean all-pole spectrum on a log magnitude scale with 36.1-Hz resolution.

A new spectral section or frame was calculated at 10-ms intervals. The 25-ms Hamming windows used in this analysis have an effective duration of about 12.5-ms. Thus the 10-ms update interval produces some spectral overlap between frames without severe oversampling. These temporal parameters are a little bit rough to preserve the onsets of formant transitions as acculately as possible. Fig. 2 shows window position.

In these running spectral analysis, the Hamming window was positioned visually by the experimenter in such a way that the portion of the burst emcompassed by the frame, had an effective duration of about 5-ms. Therefore the first frame can be said to display spectral energy from the release burst only. In contrast, Blumstein and Stevens say that a rather longer window duration is appropriate. Twenty-six coefficients were used to calculate frames regardless of consonantal voicing or CV-boundary vocalization. Although, voiceless frames and consonantal frames need fewer coefficients to specify the spectrum adequately, according to Markel and Gray (1976), adding extra coefficients gave rippled peaks closer to Fourier spectrum in detail, probably because of zeros. Therefore, in this analysis, several more coefficients were used in analysing the consonantal frames. A running spectral display was then saved in a file for the first 5 frames—or 40-ms interval—of each CV.

D. Optimal Classification Rule Based on Canonical Variable

There are two different approaches of acoustic-phonetic features processing for automatic recognition. The first models the acoustic characteristics of phonemes explicitly, using specially designed rules and features, the second models the knowledge implicitly by, for instance, representing all phonemes as probabilistic sequences of spectra. The former is called rule-based or feature-based approach. The other defines a template model or a probabilistic model.

The advantage of the first approach, the heuristic approach, is that one can incorporate various kinds of knowledge, e.g. in terms of distinctive features, relationship between phoneme class and features, and give a heuristic weight to each of several alternatives; but the lack of a methematical formalism is a severe drawback.

The advantage of the second approach, the minimum distance classifier or the minimum error probability classifier, is the possibility of a wide range of application to phonemes or pseudo phonemes, using a dynamic programming algorithm, and probabilistic scores to find the optimal sequence. Perhaps the most important aspect of this method is that the training of the probabilistic network is automatic. This allows the system to utilize very large databases.

Suppose that there are g-groups, and that observation $\mathbf{x}$ is a k-vector (for example a k-parameter frame spectrum). In order to classify consonants according

to three or more place of articulation, an optimal classification rule for multiple-groups can be applied. A set of linear classification functions is developed out of a number of measurement variables. A method using canonical variables that is a direct generalization of Fisher's approach is suitable for this application.

This method develops canonical variables based on the between-group (**B**) and within-group (**W**) covariance matrix. Fisher suggested finding the linear compound, $\boldsymbol{\lambda}$, which maximized

$$\gamma=\frac{\boldsymbol{\lambda}'\mathbf{B}\boldsymbol{\lambda}}{\boldsymbol{\lambda}'\mathbf{W}\boldsymbol{\lambda}} \tag{1}$$

Differenciating γ by $\boldsymbol{\lambda}$, and equating this to zero, the following equation is yielded

$$(\mathbf{B}-\gamma\mathbf{W})\boldsymbol{\lambda}=\mathbf{0} \tag{2}$$

This equation has a nontribial solution only if

$$|\mathbf{B}-\gamma\mathbf{W}|=0 \tag{3}$$

The solutions to this equation are eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$. There are no more than min(g-1, k) nonzero solutions.

The corresponding eigenvalues are the linear compounds $\boldsymbol{\lambda}$ that are used for discriminating. If r vectors are used, the rule becomes: Assign to Π_i if

$$\sum_{l=1}^{r}[\boldsymbol{\lambda}'_l(\mathbf{x}-\boldsymbol{\mu}_i)]^2=\min_j\sum_{l=1}^{r}[\boldsymbol{\lambda}'_l(\mathbf{x}-\boldsymbol{\mu}_j)]^2 \tag{4}$$

Since k is quite large compared to g, a convenient representation of the information results from this canonical-vector approach.

E. Feature Extraction and Data Representation

With spectrally analyzed speech, several possible representations of the frequency-by-amplitude dimensions can be chosen. Linear prediction spectra are typically represented on a linear frequency scale. In the auditory system, however, frequency on the basilar membrane is equally distributed in approximately bark intervals, which is often approximated by a simple log-frequency scale. Thus, for research employing auditory filters, a bark frequency or modified log scale (technical Mel) is probably more appropriate for representing frequency than is a linear scale. And spectral sections can be best represented in a log frequency scale and in decibels for amplitude dimensions.

Another property of running spectral dimension is the representation of time: we know that the auditory system can closely track time variations in waveforms in terms of synchrony of discharge firings with the input signal. Apparently, the important acoustic distinctions in speech vary much more slowly than the temporal processing capabilities of the ear. Therefore the limits of the representation of the time dimension for processing speech spectra should be set according to the observed rates of change in the speech signal. For speech, this limit would be placed somewhere between 1 and 20-ms.

The time intervals between spectra currently employed by different investigators are in the 5- to 10-ms range. For consonantal analysis, much finer waveform changes

have to be tracked, therefore, a frame interval shorter than 10-ms may be necessary in order to analyze, for example, bilabial stops. Since we placed the first analysis window center at the burst point by visual examination, thus incorporating a shorter portion of burst, we regarded the 10-ms frame update interval as appropriate even for bilabial stops. But in some cases 3-5-ms intervals may be more appropriate.

The span of analyzed interval is 40-ms or 5 frames, due to tractable size of information in the following statistical analysis, however, 40-ms interval seems to be sufficient for consonantal features from other studies.

When we discuss feature selection for classifying two or more distributions, we will allow a more general class of transformations. This is because the class separability, for example the probability of error due to the Bayes classifier, is invariant under any nonsingular transformations as far as classification is concerned. Feature selection is generally considered a process of mapping the original measurement into more effective features.

Therefore, feature selection for maximizing γ in equation (1) means finding a subsequence for a given m, such that the eigenvalues λ_i in the subspace are larger than those of other m-dimensional subspaces.

A set of features was selected in the following three steps.

(1) Intuitive reduction following the concept of critical bandwidth.

(2) Ordering according to the amount of information in each variable concerning the separability between groups. However, to do this, not only the individual significance level but also the joint significance level must be taken into account. Automatic stepwise entry or deletion of variables is used in BMDP7M (Dixon et al., 1977).

(3) With a linear transformation, n-dimensional feature space is reduced to a subspace of m-dimension maximizing a criterion by the process for the eigenvalue computation from which canonical vectors were obtained as described in I-D.

Features are selected intuitively from discrete Fourier spectrum, then a separability maximizing criteria is applied to reduce into minimum dimentional feature space, and then optimal transformation is applied.

II. Experiment

A. Discriminant Analysis for a Time-Window

Spectrum is computed every 10-ms after release of burst up to 5 frames. Each spectrum at a given delay after the burst can be contrasted with other consonants to distinguish consonantal features. The results show the location and distribution of stop features after burst. The features are extracted from the spectrum data set of the same order frames by the procedure described above adjusted so as to maximize the separability of consonants, and they are stored for later analysis as a set of canonical vectors.

Table 1. Frame-wise discrimination and discrimination by canonical variables for 5-frames in 3-group case.

frame	p	t	k	total	c_1	c_2	b	d	g	total	c_1	c_2
1	79.9	69.0	68.6	72.5	1	6	81.0	71.9	65.2	72.8	1	5
2	78.5	64.1	80.0	74.2	5	3	84.8	76.2	80.8	80.8	2	3
3	72.7	52.8	81.4	68.9	2		71.4	73.1	73.2	72.5	4	6
4	66.4	47.2	65.5	59.7			59.7	66.2	68.6	64.8		
5	61.2	47.9	57.9	55.7		4	54.8	61.9	66.2	60.9		
c-all	86.5	70.7	87.6	81.6			89.0	82.7	82.6	84.8		

Table 2. Frame-wise discrimination and discrimination by canonical variables for 5-frames in 4-group case.

frame	p	t	$\tilde{k}$	$\dot{k}$	total	c_1	c_2	c_3	b	d	$\tilde{g}$	$\dot{g}$	total	c_1	c_2	c_3
1	81.7	72.8	87.9	74.1	78.9	1	3		83.1	77.7	84.2	69.8	79.8	1	3	7
2	79.6	70.7	90.8	75.0	78.3	9	4	2	82.8	82.3	80.1	76.7	81.2	2	4	6
3	61.6	49.3	79.3	81.0	63.6	6	7		67.9	70.8	77.2	74.1	71.6			9
4	66.4	50.0	66.7	64.7	60.8				54.1	68.1	69.6	66.4	63.3		5	
5	52.9	46.9	55.2	56.0	51.8	8		5	52.4	62.3	65.5	68.1	60.3	8		
c-all	87.5	76.6	90.8	84.5	84.1				87.9	86.9	90.6	76.7	86.6			

Analysis results are shown in Tables 1 and 2 for 3-group and 4-group discrimination respectively. In 4-group analysis, velars were divided into ones followed by a back-vowel and another palatalized ones followed by a front-vowel.

(1) Four-group analysis resulted in higher correct discrimination rate than 3-groups.

(2) Allover correct discrimination decreases as the analysis frame moves toward vowel from consonant plosion. This means more stops feature reside around the burst portion.

(3) The second frame analysis (10-ms after burst) achieved the highest result.

Therefore, the 10-20-ms waveform after burst possesses most important stop features. The shape of time window is one important factor for stop feature extraction, that is, what part of waveform have to be integrated into a spectrum. The first frame discrimination obtained here, however, was inferior to previous discrimination described in Kitazawa et al. (1982). The reason is that the present analysis uses a 25-ms Hamming window, while the previous analysis used a half-Hamming window like the one suggested by Blumstein and Stevens (1978).

Each analysis frame discriminated consonants differently. Bilabials /p/ and /b/ tended to be discriminated best in the first frame, velars /$\tilde{k}$/ and /$\tilde{g}$/ in the second, dentals /t/ in the first, /d/ in the second, and palatalvelars /$\dot{k}$/ and /$\dot{g}$/ in the second and third. These results reflect location of each phoneme feature, though, there is no direct correspondence with recognition rate. Since this feature was evaluated by between-phoneme discrimination, the discrimination rate is method dependent and individual phoneme discrimination rate is in reciprocal relationship.

As a consequence the integrated features from multiple frames can be expected to improve the rate of correct classification to some extent.

B. Discrimination with a Set of Canonical Variables

Once the presence or absence, or degree of presence or absence, of each feature of a pattern has been evaluated, all this information must be combined in order to determine the overall goodness of the match to the phoneme pattern.

There should be a physiological background of integration of features, however, currently there seems to be no established hypothesis, therefore, the integration procedure used here is a purely mathematical one, that is, a statistical optimal decision process.

The consonantal discriminant features were represented in compact canonical variables for each frame as described in II-A. Canonical variables of a frame comprise two (in 3-groups) or three (in 4-groups) variables which were put together to compose 10 (in 3-groups) or 15 (in 4-groups) input variables over all frames. Although, it may be apparent that more features given a more accurate result in recognition, it is recommended to select variables by the stepwise enter and remove process in practical application. Selected variables are shown in Table 1 and 2 in the c column. In both voiced and voiceless case, individual consonant discrimination rate became better than the best one among 5 frames, and overall score was improved as well. This shows the features for place of articulation are effectively integrated.

C. Maldistribution of Phonetic Features

In the stepwise variable selection process of the discriminant analysis, some effective features were evaluated from a canonical variables set. This selection reflects which canonical variable is most effective. Consonantal information does not distribute uniformly among frames, but rather some frames are redundant or highly correlated with others, eventhough they are individually effective, because they are coarticulated under the adjacent phoneme influence. Therefore each variable is effective for discrimination to the extent it is relatively independent.

The resultant selection of variables is shown in Table 1 and 2 in the row Ci ($i=1, 2,$ or $1, \cdots\cdots, 3$). The first canonical variable is for bilabial and velar distinction and the second is for dental and velar distinction. Both 3- and 4-group discrimination selected the canonical variables in first and second frames with high significance values for classification. On the other hand, third, fourth, and fifth frames were not so effective, especially, the fourth frame was almost redundant or unnecessary to compute.

From these results, we can consider non-uniform spectrum observation is appropriate, the burst spectrum have to be observed with a small frame increment, but CV-boundary spectra are sufficient to be observed with larger increment. As can

be seen from the table, even the fifth frame, 40-ms from the burst, is still effectively recognizable, possibly due to difference in duration of aspiration interval as shown in histograms of VOT for each consonant (Fig. 1). Since for some velar samples the onset of vowel was very slow, late spectrum may still be effective for discrimination. The short-time spectra used in this study represent vowel independent spectral features only, and do not represent time-varying features such as formant transitions or formant locus which are dependent on the following vowel. Moreover, it should be noted that time-varying energy feature was not used in this experiment, because spectra used were power normalized.

III. DISCUSSION

Some authors view discriminant analysis as a technique for the description and testing of between-group differences. Although the stepwise variable selection process selects some significant variables, the best set of two variables may not include the best single variable. The observation of selected variables in the different frames, whether consonant is voiced or voiceless, or whether the data-base is supposed to discriminate difference between 3- or 4-groups, may give us some insight into stop consonant features. The resultant selections of variables are shown in each case in the column in Table 3. The number in the column represents the significance order under different discriminant conditions. Every frame selects some ten variables out of 28 variables according to the between-group F-values which reflect the extent of contribution of each variable. Generally, high frequency components are effective near the burst point, and decrease afterward as time goes by.

Differences of features between voiceless and voiced consonants can be stated as follows:

For voiceless stop discrimination, low frequency components around 200-Hz are primary features, while for voiced stop discrimination, middle frequency components around 2,600-Hz are primary features. The possible reason is that voiced plosives are superimposed with low-frequency energy that was caused by early voicing before plosion. So, the spectrum of voiced consonants around low-frequency is relatively similar regardless of place of articulation, and higher frequency components thus are the first selected. On other hand, in voiceless bilabials dominates low frequency components, apparently contrasting from the other place of articulation.

The fact that the performance reported here is better than those reported in competing works may have little significance because they have been obtained with different data sets. The true important result is that performances improve remarkably as more significant features are integrated into the rules. A wise integration of powerful features is more important than a refinement of an algorithm working on sets of data which have not been carefully selected. The good performance achieved so far makes this approach suitable to be used in a general purpose speaker independ-

Table 3. Resultant variable selection for each frame. Variables are ordered according to the significance for discrimination.

Spectrum variables			4-group										3-group									
			voiceless					voiced					voiceless					voiced				
channel	kHz	frame	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	.03		11		13	6		4							1	1		8				
2	.12		1	1	11				8	4	5		1	2							7	8
3	.23					7					4			10	12						3	6
4	.34						11		1	14		8			17		8		6		15	
5	.45				7			14			6	9		1	13			14	15	11	10	11
6	.56								3				6									10
7	.66					9					9					6	6		14	13	9	9
8	.77		2		4	2			7	1	3		2		6				16	7	4	
9	.88			11			1	13	16			2				4	1	10	10			3
10	.99						3	9		10	13	5			15	9	3	13		2	8	7
11	1.12			9	3		2		15	7		4		7	2	10	4			16		
12	1.26								17		2										2	2
13	1.42		3	10			7	2			11		3				5	2		12	6	4
14	1.62				12				4		14	3			14	7			2	17	11	12
15	1.84		6	5	5			7			12	6	9	3	9	12		4				
16	2.07						5		13	3	7	7							7	3	5	5
17	2.35		5	2	10	5		11	10	6	8	10	4	4	7	11	10	12	5	9	14	
18	2.65			12	1	4	8	8	2	2	1	1		11	4	2	2	9	1	1	1	1
19	3.02							10	14	12	15	12	5	5	5			5	3			
20	3.41		10	3	2	1	6		9	9	10			8	8	3	7			6	12	14
21	3.85		8	8	9			1	6					9	11			1	8	15		
22	4.34		9	13		3	4	6		5			7			8	9	7	12	4		
23	4.90			6	8			3	5					6	3	5	11	6	4	5	13	13
24	5.53		7		14			12	12			11	8	12	16			11	11			
25	6.23			7	6		9		11	13				13	10				13	8		
26	7.03							15						14						10		
27	7.92		4	4			10	5		8			10					3	9	14		
28	8.80		12			8				11			11									

ent speecch recognition system. Also, these results partially support the acoustic-invariance hypothesis, because the performance was achieved under vowel context and speaker independent condition.

Several problems still remain to be investigated further. Present result is only preliminary. For example, time varying features may be important. Though implicitly included in the current result, they are not yet explicitly extracted.

References

Blumstein, S.E., and Stevens, K.N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust.

Soc. Am. 66, 1001-1017.

Blumstein, S.E., and Stevens, K.N. (1980). "Perceptual invariance and onset spectra for stop consonant in different vowel environments," J. Acoust. Soc. Am. 67, 648-662.

Delattre, P.C., Liberman, A.M., and Looper, F.S. (1955). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. 27, 769-773.

Dixon, W.J., and Brown, M.B. (1977). *BMDP-77*, University of California Press.

Ide, K., Makino, S., and Kido, K. (1983). "Recognition of unvoiced plosives using Time Spectram Pattern," J. Acoust. Soc. Japan 39, 321-329.

Kewley-Port, D. (1980). "Representation of spectral change as cues to place of articulation in stop consonants," Res. Speech Percept. Tech. Rep. No.3, Indiana University.

Kewley-Port, D. (1983). "Time-varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am. 73, 322-335.

Kitazawa, S., and Doshita, S. (1982). "Discriminant analysis of burst spectrum for Japanese initial voiceless stops," Studia Phonologica, XVI, 48-70.

Lachenbruch, P.A. (1975). *Discriminant Analysis*, Hafner Press, New York.

Lisker, L., and Abramson, A.S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," Word 20, 384-422.

Markel, J.D., and Gray Jr., A.H. (1976). *Linear Prediction of Speech*, Springer-Verlag.

Searle, C.L., Jacobson, J.Z., and Rayment, S.G. (1979). "Stop consonant discrimination based on human audition," J. Acount. Soc. Am. 65, 799-809.

Searle, C.L., Jacobson, J.Z., and Kimberly, B.P. (1980). "Speech as patterns in the 3-space of time and frequency," in *Perception and Production of Fluent Speech*, edited by R.A. Cole (Erlbaum, Hillsdale, NJ), 73-102.

Stevens, K.N., and Blumstein, S.E. (1978). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. 64, 1358-1368.

Winitz, H., Scheib, M.E., and Reeds, J.A. (1972). "Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech," J. Acoust. Soc. Am. 51, 1309-1317.

(Aug. 31, 1983, received)